SECTION 4. COMPUTER SCIENCE

4.1 Implementation of information and mathematical models and algorithms in the software environment

The main component of object-oriented analysis in the development of an information system is the decomposition of the problem into separate classes of concepts (conceptual classes) or objects. Domain modeling is one of the initial stages of system design, which is required to identify, classify, and formalize information about all aspects of the domain that determine the properties of the system being developed.

The description of the subject area of epidemiological system research was carried out using the UML language and the Visual Paradigm CE tool integrated in the Netbeans IDE. When modeling the subject area, it was assumed that the designed software environment is intended for entering, saving and processing information about epidemiological systematic studies [44].

The system must provide the following basic functions:

- collection and processing of statistical epidemic information;
- development of compartmental mathematical models based on statistical data;
- quantitative forecasting of the development of the spread of epidemic diseases;
- development of immunoprophylactic measures;
- making decisions about the shape of the epidemic curve.

The main functions actually describe the technical task of this system.

The characters. The selection of categories of system users is, as a rule, carried out informally. When choosing actors, we used the features proposed in:

- users participate in various (independent) business processes;
- users have different rights to perform actions and access information;
- users interact with the system in different modes: from case to case, regularly, constantly.

Based on these characteristics, we come to the following categories of users in a first approximation:

- epidemiologists – scientists who work with epidemiological models and statistical data, develop diagrams of transitional states;

- system analysts – scientists who develop epidemiological mathematical models, carry out their quantitative and qualitative research;

- sanitary-epidemiological service (SES) - SES specialists working on collection, storage and primary processing of statistical data on the spread of epidemic diseases. Forming a representation of the diagram of the use of the software environment of epidemiological system studies. An epidemiologist is called an Epidemiologist, a systems analyst is called a Systems Analyst, and a SES specialist is called an Epidemic Service Officer [45].

Options for use. The analysis of the technical task allows us to identify the following options for use:

- collection of statistical epidemic information;
- processing of statistical epidemic information;
- development of compartmental mathematical models;
- quantitative forecasting of the development of the spread of epidemic diseases;
- development of immunoprophylactic measures;
- making decisions about the shape of the epidemic curve [46].

Relationships in the usage chart. In the first approximation, two types of relationships are used in the usage diagram: the association between the actor and the use case, which shows how the actor in one way or another interacts (gives input, receives a result) with the use case; dependence between use cases, which shows that one use case depends on another (in this case, the dependence is that in a certain scenario of an independent use case, a dependent use case scenario can be inserted as a subsequence of actions in a certain place - the "extend" stereotype).

Based on how actors participate in scenarios belonging to certain use cases, it is advisable to establish associations:

- for epidemiologists with options for use "development of compartmental mathematical models", "quantitative forecasting of the development of the spread of epidemic diseases", "development of immunoprophylactic measures", "making decisions about the shape of the epidemic curve";

- for system analysts with options for use "development of compartmental mathematical models", "quantitative forecasting of the development of the spread of epidemic diseases", "development of immunoprophylactic measures", "making decisions about the shape of the epidemic curve";

- for specialists of SES with options of use "collection of statistical epidemic information", "processing of statistical epidemic information".

In addition, there are dependencies (stereotype extend) between certain usage options. Thus, the usage option "processing of statistical epidemic information" depends on "collection of statistical epidemic information"; the use case "development of compartmental mathematical models" depends on "processing of statistical epidemic information"; the use options "quantitative forecasting of the development of the spread of epidemic diseases" and "development of immunoprophylactic measures" depend on "development of compartmental mathematical models"; the use option "making decisions about the shape of the epidemic curve" depends on "quantitative forecasting of the development of the spread of epidemic diseases" [47].

Usage chart. So, if we look at this usage model from the most general point of view, we can see that the model contains: an internal modeled system in the form of a set of usage options connected by dependencies; external environment in the form of actors; the connection between the modeled system and the external environment in the form of associations between actors and use cases. The internal part, separated by boundaries, has a specific name in UML - subject [48].

Implementation of use cases. For each of the use cases, the goal was to describe all the scenarios that make up the use case. At the same time, two types of diagrams were used - action diagrams and sequence diagrams.

Thus, the use case for the collection of epidemiological information is represented by an action diagram (Fig. 1)

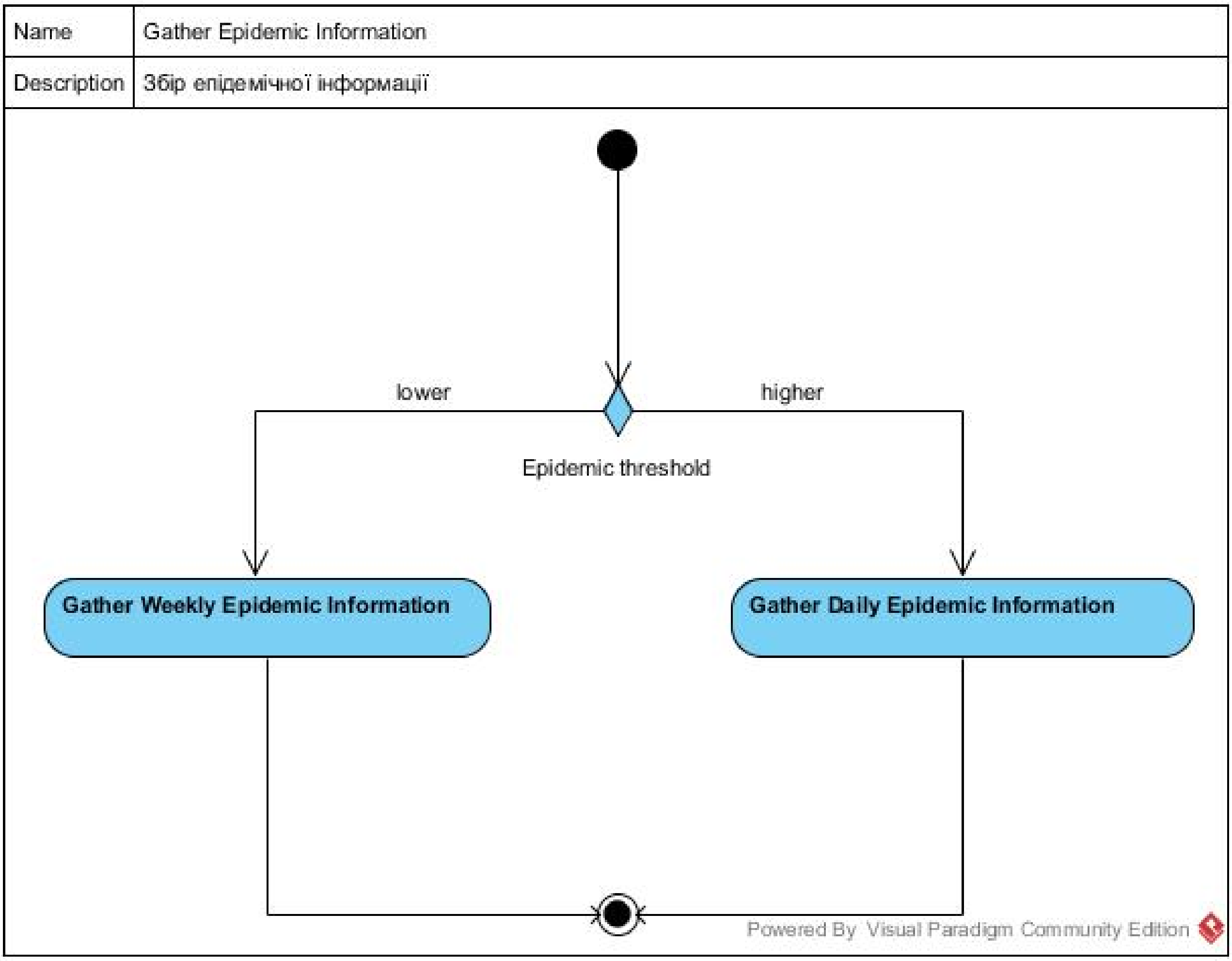
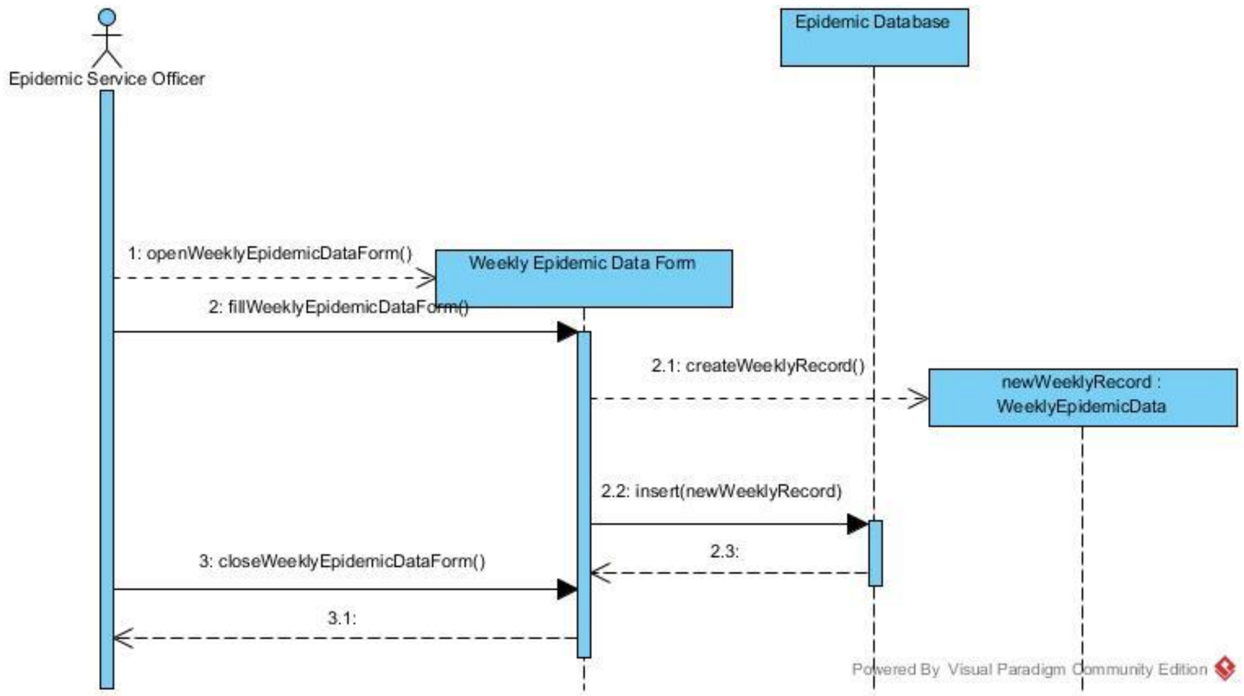


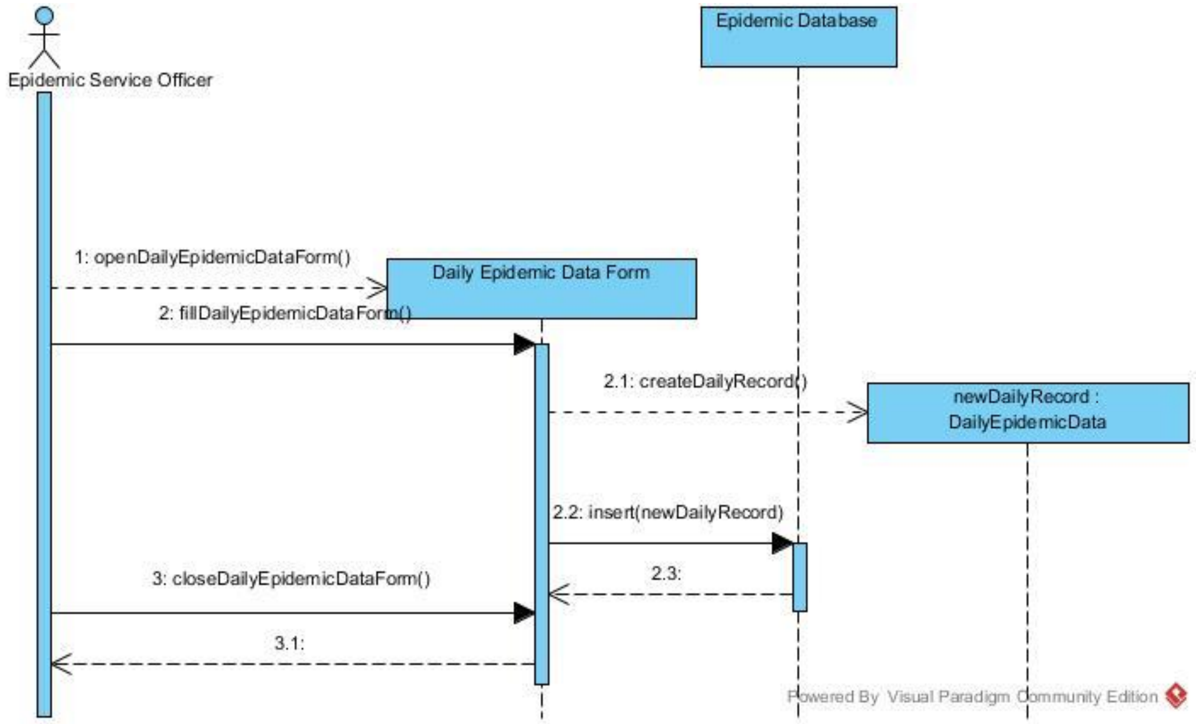
Figure 1. Action diagram "Collection of epidemic information"

The procedure for collecting epidemic information (weekly or daily data) is determined by the incidence exceeding the epidemic threshold value [49].

At the same time, the use options "Collection of weekly epidemic information" and "Collection of daily epidemic information" are implemented by sequence diagrams (Fig. 2).



a)



b)

Figure 2. Sequence diagrams for the use cases "collection of weekly epidemic information" (a) and "collection of daily epidemic information" (b)

The use case "processing of statistical epidemic information" includes the calculation of epidemic thresholds based on previously accumulated information in the database and the construction of epidemic curves (Fig. 3)

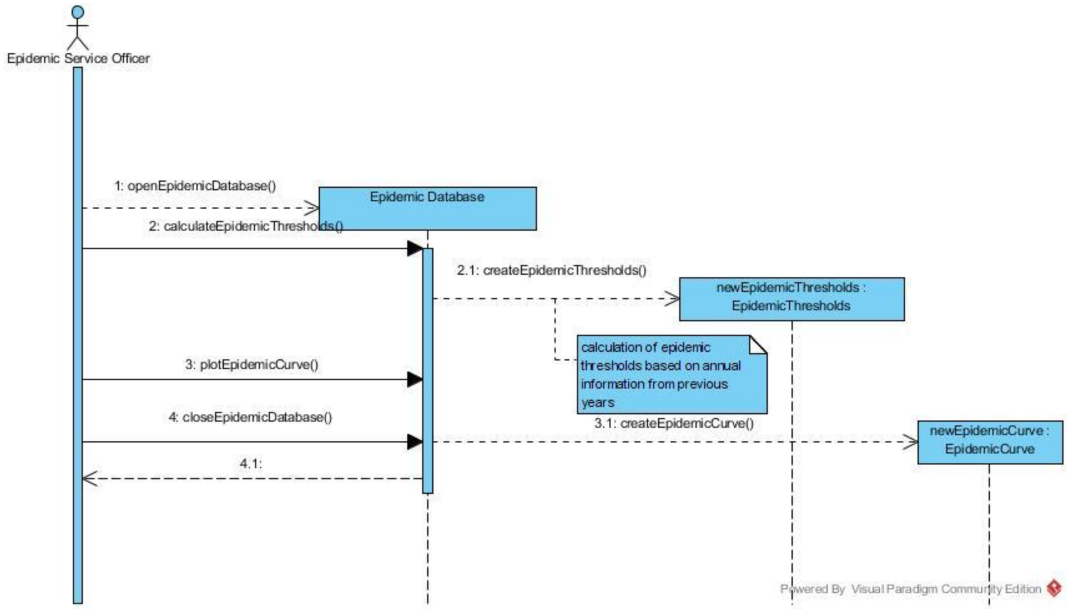


Figure 3. Sequence diagram "processing of statistical epidemic information"

Compartment model development involves building models with one or two virus strains in mind (Fig. 4).

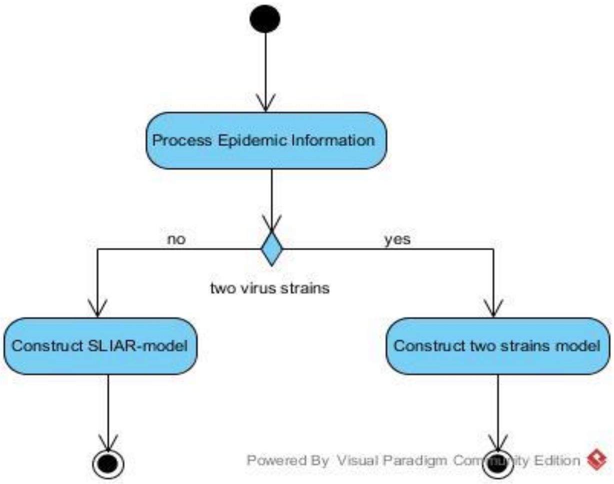


Figure 4. Activity diagram for the "compartment model development" use case

The development of immunoprophylaxis measures is based on a previously developed compartmental model. Such a model should be accompanied by controls such as the intensity of immunoprophylactic measures and corresponding quality criteria (Fig. 5)

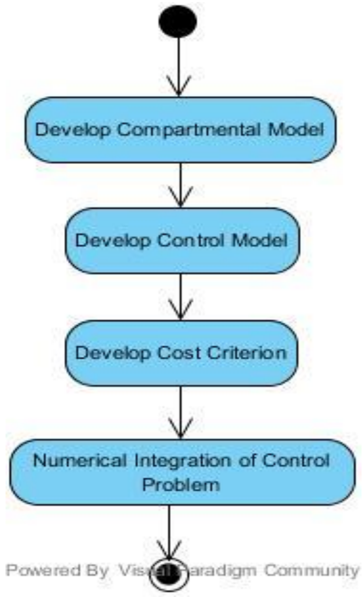


Figure 5. Action diagram for the development of immunoprophylactic measures

Quantitative forecasting based on a previously developed compartmental model begins with the identification of model parameters. Next, the numerical integration algorithm of the model equations is applied (Fig. 6).

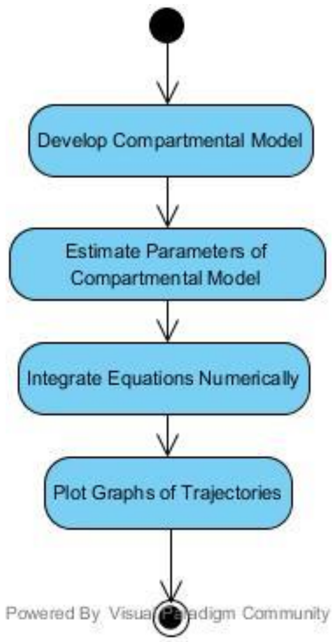
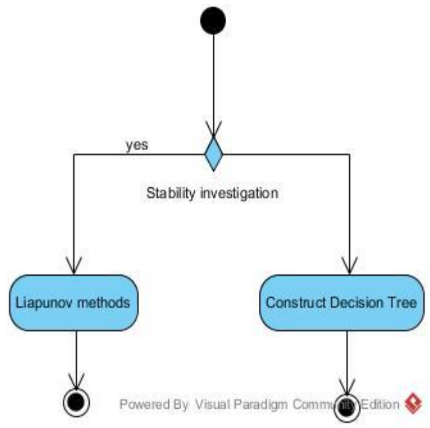
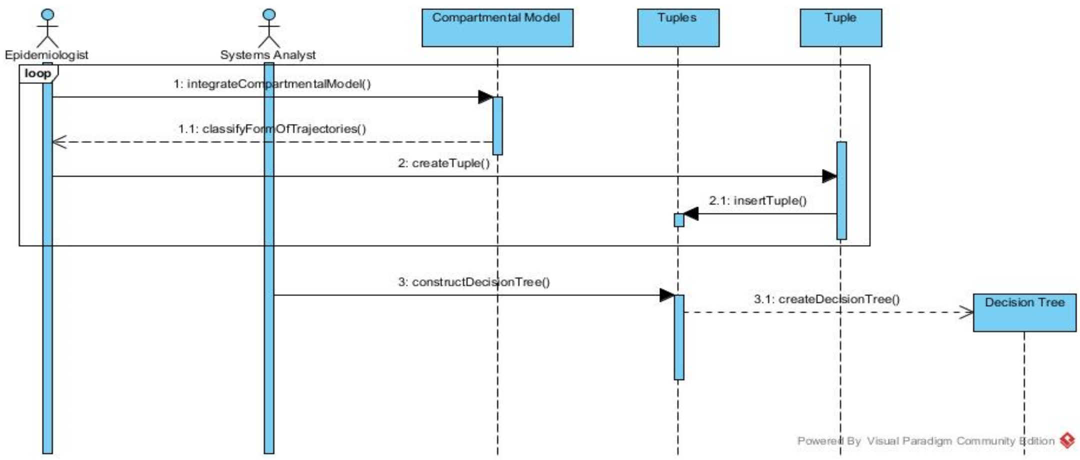


Figure 6. Action diagram for quantitative forecasting.

Qualitative forecasting of the epidemic curve includes both the study of the stability of endemic states of models using Lyapunov methods, and the analysis of more complex forms of trajectories (such as the prediction of repeated outbreaks) - using data mining technology, through the construction of a decision tree (Fig. 7)



a)



b)

Figure 7. a) action diagram for qualitative forecasting of the epidemic curve; b) sequence diagram for building a decision tree.

The developed software environment is actually a decision support information system intended for operation in the conditions of a research laboratory of infectious diseases. An important task is forecasting the development of the epidemic with the prediction of repeated outbreaks, as well as the development of optimal schemes of preventive measures. The previously built information model was used in the development of a decision support system for forecasting and development of optimal prevention schemes for the influenza epidemic at the regional level [50].

The input data entering the regional SES database is provided by the district (city) state sanitary-epidemiological supervision weekly (in the period of rising incidence, the data is transmitted daily). In addition to data for the construction of epidemic curves, the database contains arrays of data with the results of simulation simulation of epidemics and qualitative conclusions of epidemiologists, which will be used for the construction of knowledge structures.

It is possible to increase the efficiency of prevention of acute respiratory infections by applying methods of intelligent data analysis. It is appropriate here to use data mining tools, among which the decision tree induction method and the sequential coverage algorithm are chosen. To solve the problems, the use of these methods in the environment of operational analytical processing is the most productive. Based on the accumulated information, a decision-making support block for intelligent data analysis has been developed, which for the user is represented by a set of forms for the development of models, their quantitative and qualitative analysis [51].

In this case, the subsystem of intelligent data analysis is implemented on the basis of developed compartmental models of the forecast of the development of the acute respiratory syndrome epidemic. Therefore, a necessary condition for the analysis is the formation of analytical metadata for the processing system.

To assess the sensitivity and specificity of the built models, a cross-validation test was used, i.e. a method of evaluating the model and its behavior on independent data. To determine the classification value of the resulting prognostic model, the ROC curve was used, followed by the determination of the area under it [52].

For the convenient application of the constructed models and working with the database, in order to make a decision on epidemic forecasting and preventive measures, a user application has been developed, which contains forms for filling in and displaying information about the dynamics of morbidity and calculating the forecast of the development of the epidemic and its qualitative analysis. If we consider the system from the user's point of view, the most important criteria for him when working with data are work with a user-friendly interface, high data processing speed, and the ability to customize the system to his own needs.